

Weekly Report

Lu Junhua

2016 年 10 月 17 日

Done

- Spent much time in editing the video. Added captions and sound to the video, and modified several times according to others' suggestions.
- Discussed and replied to Prof. Peng's questions.

To do

- Prepare for the group meeting.
- Assign fundamental tasks to the students of our project group.
- Go to America.

论文

VAST 16 *Patterns and Sequences: Interactive Exploration of Clickstreams to Understand Common Visitor Paths* 本文虽然被选入tvcg但是文章所描述的可视设计比较一般、用户调研也没有明确可信的。主要针对的是日志数据, 希望进行一些时序数据模式的挖掘。文章对其所在进行这个项目中的步骤说的十分详细也十分诚实, 说自己曾经失败了多少种数据处理算法等等, 我估计是这个打动了reviewer.

ICDM 15 *A General Suspiciousness Metric for Dense Blocks in Multimodal Data* 对于多模态数据(其实可以认为是多维数据) 中异常行为检测的方法。不同于之前看的基于关联规则的各种方法, 本文用的是从张量中提取dense block 的方法。文章先指出寻找dense block的必要性: 为什么与异常行为有关联。然后定义了一些metric并发现之前的一些算法都无法满足所有这些条件, 最后提出了自己的一套方案名为CrossSpot。之前未曾仔细了解过张量相关知识, 从本文以及网上搜索到来的信息来看, 其应用非常广泛(并且和之前看到的一些矩阵相关的方法有不同之处), 后面可能还做相关内容的调研。

VAST 16 *NameClarifier: A Visual Analytics System for Author Name Disambiguation* 本文是介绍一种利用可视化方法解决论文中人名重名的问题。一方面, 设计十分简洁, 清晰, 配色温和; 另一方面很好的满足了可视分析中将人带入的要求。根据数据库已有的确认的论文重名数据, 对未确认的数据进行交互式分析, 定义一些相似性度量等等。文章本身没有用非常复杂的模型或者算法, 因为作者认为之前那些去重名的挖掘算法都太广了而且完全黑盒, 效果不佳。在设计问题时作者考虑的问题很全, 包含了对数据库已确认数据也进行准确度估计等等。当然这里其实对于人的带入成分比以往的系统要高, 有些东西不一定完全是符合ground truth, 作者也不可能一一去全部核实。因此作者最后也写到希望能加入一些更好的算法来

提高效率.

VAST 16 *DropoutSeer: Visualizing Learning Patterns in Massive Open Online Courses for Dropout Reasoning and Prediction* 文章是对MOOC中退学率的一个预测, 对于为什么参加MOOC的会中途退出等原因探索的一个系统. 这个系统与去年的egoSlider及其神似, 包括可视设计方面, 个人认为在case study中也没有很好的体现其预测性的功能, 可能仍然是说理为主. 当然这个数据还是比较新颖, 可视设计上面小巧的glyph我们也确实模仿不来或者说不知道该怎么用好它, 这一块也是后面要多多学习的, 一方面时序数据常用一方面也确实比较美观.